

[illegible]

OS Homo sapiens.
XX
PN MO9928344-A2.
XX
PD 10-JUN-1999.
XX
PF 30-NOV-1998; 98MO-EP07714.
XX
PR 09-APR-1998; 98EP-0870078.
XX
PR 28-NOV-1997; 97EP-0870195.
XX
PA (INNO-) INNOGENETICS NV.
XX
PI Meheus L, Raymackers J, Union A;
XX
DR WPI; 1999-385357/32.
XX
PT New peptide derived from intermediate filament proteins
XX
PS Example 1; Fig 2; 73pp; English.
XX

AAAY2954-57 represent amino acid sequences of human filaggrin clones. The
CC specification describes peptides derived from any variant of natural
CC filaggrin or any variant of intermediate filament proteins. These
CC peptides contain at least one citrulline residue which is crucial
CC for reacting with antibodies that are present in sera from patients
CC with rheumatoid arthritis. The peptides constitute immunogenic
CC determinants of antibodies present in patients with rheumatoid
CC arthritis. The peptides, antibodies, immunotoxins and intermediate
CC filament proteins can be used for the preparation of a therapeutic or
CC of a diagnostic for rheumatoid arthritis. The peptides can also be
CC used for identifying compounds which modulate the interaction between
CC an autoantigen and a rheumatoid arthritis specific autoantibody. The
CC products can also be used for the diagnosis and treatment of other
CC autoimmune diseases e.g. systemic lupus erythematosus, discoid lupus
CC erythematosus, scleroderma, dermatomyositis, or Sjogrens syndrome.

Sequence 330 AA;

PF 08-JUN-1990; 90US-0143578.
XX
PR 08-JUN-1990; 90US-0535206.
XX
PA (USSH) NAT INST OF HEALTH.
XX
PI Ortaldo J, Young H, Anderson S;
XX
DR WPI; 1991-245694/33.
DR N-PSDB; AAQ13115.
XX
PT DNA encoding a natural killer cell receptor - used to develop
PT prods. for the immuno-detection and immuno-therapy of tumours
XX
PS Disclosure; Fig 2; 30pp; English.
XX
XX Overlapping clones, which make up the cDNA sequence from which this
CC sequence was deduced, were isolated from a murine PBL cDNA library
CC prepared in lambda gt10. The purified protein can mediate the
CC cytolytic activity of mammalian cells. It specifically distinguishes
CC tumour cells making it a candidate for the development of products
CC for the immunodetection and immunotherapy of tumours.
CC See also AAQ13114.
XX
XX Sequence 982 AA;

	Query Match	29.2%	Score 73.5;	DB 12;	length 982;
	Best Local Similarity	39.3%	Pred. No.	1.5'	
Dbs	Matches 22:	Conservative	6;	Mismatches	21; Indels 7; Gaps 2;
Oy	1 STGHGSGCHSHTTTTCGRS--- <td>:</td> <td>: </td> <td> </td> <td> </td>	:	:		
	: :: :	:	::		:
Bd	693 ssrggssrstsrsstsrtsrylstrslrstlsrsparsrhspnkysdgshhs 748	:	:		

Query Match	82.1%	Score 207;	DB 20;	Length 330;
Best Local Similarity	81.2%	Pred. No. 5	1e-17;	
Matches	39;	Conservative	4;	Mismatches 5; Indels 0;
OY	2	TGHSQSHHTTQGRDASRRSSGSRSTSRRTROEDSGDGSRRSGS	49	
Db	75	sghsqshhttsctsggrsdasngsrsasrttrndsgsgdgsrrnsgs	122	

RESULT	8
AAM21759	
ID	AAM21759 standard; protein; 63 AA.
XX	
AC	AAM21759;
DT	12-OCT-2001 (first entry)
XX	
DE	Peptide #8193 encoded by probe for measuring cervical gene expression
XX	
TM	Probe; human; microarray; gene expression; cervical epithelial cell;

	RESULT	7
AAAR13320	ID	AAAR13320 standard; Protein; 982 AA.
XX	AC	AAAR13320;
XX	DJ	22-OCT-1991 (first entry)
XX	DE	Murine Natural Killer receptor.
XX	KM	NK; cytotoxic drugs; tumour cell.
XX	OS	Homo sapiens.
XX	FH	Key
FT	FT	Protein
FT	FT	Peptide.
XX	FT	/label= mature_NK_receptor 1..7
XX	PN	/label= partial_signal_peptide
XX	PD	US7535206-A.
XX		09-JUL-1991.

OS	Homo sapiens.
XX	
PN	MO200157278-A2.
XX	
PD	09-AUG-2001.
XX	
PE	30-JAN-2001; 2001WO-US00670.
XX	
PR	04-FEB-2000; 2000US-0180312.
PR	26-MAY-2000; 2000US-0207456.
PR	30-JUN-2000; 2000US-0608408.
PR	03-AUG-2000; 2000US-0612366.
PR	21-SEP-2000; 2000US-0234687.
PR	27-SEP-2000; 2000US-0236359.
PR	04-OCT-2000; 2000GB-0024263.
XX	
PA	(MOLE-) MOLECULAR DYNAMICS INC.
XX	
PI	Peun SG, Hanzel DK, Chen W, Rank DR;
XX	
DR	WPI; 2001-488901/53.
XX	
PT	Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human cervical epithelial cells -

XX Claim 27; SEQ ID NO 26585; 487pp; English.

XX PS

CC The present invention relates to human single exon nucleic acid probes (SENPs: see AA110068-AA128459). The present sequence is a peptide encoded by one such probe. The SENPs are derived from human HeLa cells. The SENPs can be used to produce a single exon microarray, which can be used for measuring human gene expression in a sample derived from human cervical epithelial cells. By measuring gene expression, the probes are therefore useful in grading and/or staging of diseases of the cervix, notably cervical cancer.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pub/published_pcr_sequences](http://wipo.int/pub/published_pcr_sequences).

CC XX

SO Sequence 63 AA:

Query Match 28.8%; Score 72.5; DB 22; Length 63;
Best Local Similarity 43.1%; Pred. No. 0.091;
Matches 22; Conservative 4; Mismatches 20; Indels 5; Gaps 2;

QY 1 STGHSGSHSHTTTQGRSDASRGSSGSRSTSRRTDQSGDGS---RH 46
|| ||| | : | | | | ||| | : || | ||
5 stsssgststgsssgssssasrsgrsstsrrsssssgspsprrrh 55

Db

RESULT 9
AAM38081
ID AAM38081 standard; Protein; 63 AA.

XX AC AAM38081;

XX DT 17-OCT-2001 (first entry)

XX DE Peptide #12118 encoded by probe for measuring placental gene expression.

XX KM Probe: microarray; human; placenta; antenatal diagnosis;

XX KW genetic disorder.

XX OS Homo sapiens.

XX PN MO200157272-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00663.

XX PR 04-FEB-2000; 2000US-0180312.
26-MAY-2000; 2000US-0207456.
30-JUN-2000; 2000US-0608408.
03-AUG-2000; 2000US-0632366.
21-SEP-2000; 2000US-0234687.
27-SEP-2000; 2000US-0236359.
04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX P1 Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI: 2001-488897/53.

XX PT Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human placenta -

XX PS Claim 27; SEQ ID NO 38350; 654pp; English.

XX CC The present invention relates to single exon nucleic acid probes (SENPs: see AA113115-AA157546). The present sequence is a peptide encoded by one such probe. The probes are useful for producing a microarray for predicting, measuring and displaying gene expression in samples derived from human placenta. The probes are useful for antenatal diagnosis of human genetic disorders.

XX SQ Sequence 63 AA:

Query Match 28.8%; Score 72.5; DB 22; Length 63;
Best Local Similarity 43.1%; Pred. No. 0.091;
Matches 22; Conservative 4; Mismatches 20; Indels 5; Gaps 2;

QY 1 STGHSGSHSHTTTQGRSDASRGSSGSRSTSRRTDQSGDGS---RH 46
|| ||| | : | | | | ||| | : || | ||
5 stsssgststgsssgssssasrsgrsstsrrsssssgspsprrrh 55

Db

RESULT 10
AAB93496
ID AAB93496 standard; Protein; 407 AA.

XX AC AAB93496;

XX DT 26-JUN-2001 (first entry)

XX DE Human protein sequence SEQ ID NO:12809.

XX KM Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX OS Homo sapiens.

XX PN EP1074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 99JP-0248036.
27-AUG-1999; 99JP-0300253.
11-JAN-2000; 2000JP-0118776.
02-MAY-2000; 2000JP-0183767.
09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX DR WPI: 2001-318749/34.

XX PT primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -

XX PS Claim 8; SEQ ID 12809; 2537pp + CD ROM; English.

XX CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

[illegible]

FT	/note=	"potential casein kinase II phosphorylation site"
FT	663	
FT	/note=	"potential protein kinase C phosphorylation site"
FT	666	
FT	/note=	"potential protein kinase C phosphorylation site"
FT	671	
FT	/note=	"potential protein kinase C phosphorylation site"
FT	683	
FT	/note=	"potential casein kinase II phosphorylation site"
FT	685	
FT	/note=	"potential protein kinase C phosphorylation site"
FT	687	
FT	/note=	"potential cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT	690	
FT	/note=	"potential casein kinase II phosphorylation site"
FT	694	
FT	/note=	"potential protein kinase C phosphorylation site"
FT	697	
FT	/note=	"potential protein kinase C phosphorylation site"
FT	699	
FT	/note=	"potential cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT	701	
FT	/note=	"potential protein kinase C phosphorylation site"
FT	702	
FT	/note=	"potential casein kinase II phosphorylation site"
FT	704	
FT	/note=	"potential casein kinase II phosphorylation site"
FT	704	
FT	/note=	"potential protein kinase C phosphorylation site"
FT	710	
FT	/note=	"potential protein kinase C phosphorylation site"
XX	US6162601-A.	
XX	19-DEC-2000.	
PD	08-AUG-1997;	97US-0910925.
XX	08-AUG-1997;	97US-0910925.
PR	08-AUG-1997;	97US-0910925.
XX	(INCY-) INCYTE PHARM INC.	
PA	Bandman O, Shah P, Lal P;	
PI	WPI: 2001-090399/10.	
XX	DR N-PSDB; AAF24910.	
DR	Novel nucleotide sequence encoding human pinin splice variant, useful for treating AIDS, Addison's disease, allergy, anaemia, asthma, atherosclerosis and adult respiratory distress syndrome -	
PT	Claim 1; Fig 1A-G; 38pp: English.	
PS	The present sequence represents a human pinin splice variant (PININ). PININ	
XX	CC polypeptides and polynucleotides are useful for treating, diagnosing and	
CC	preventing developmental, vesicle trafficking, neoplastic and	
CC	immunological disorders. Antagonists of PININ polypeptides are useful	
CC	for treating AIDS. Addison's disease, adult respiratory distress	
CC	syndrome, allergies, anaemia, asthma, atherosclerosis, bronchitis,	
CC	cholecycolitis, Crohn's disease, ulcerative colitis, atopic dermatitis,	
CC	demaecomystitis, diabetes mellitus, emphysema, erythema nodosum, atrophic	
CC	gastritis, glomerulonephritis, gout, Graves' disease, hyperosinophilia,	
CC	irritable bowel syndrome, lupus erythematosus, multiple sclerosis,	
CC	mysthenia gravis, myocardial or pericardial inflammation,	
CC	osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid	
CC	arthritis, scleroderma, Sjogren's syndrome, Werner syndrome, and	

CC autoimmune thyroiditis, complications of cancer, hemodialysis, and
CC extracorporeal circulation, trauma, and vital, bacterial fungal,
CC parasitic, protozoal, and helminthic infections.
XX

SQ Sequence 717 AA;

Query Match 27.4%; Score 69; DB 22; Length 717;
Best Local Similarity 38.8%; Pred. No. 3.6;
Matches 19; Conservative 6; Mismatches 22; Indels 2; Gaps 1;

OY 1 STGSGSOSHHTTTCGRSDASRGSSGSRSTSEETDQEDSGGSRHSGS 49
I: | | | : : | : | | | | | : | | |
Db ssssssssttsssgsssgsssgssrrss--sssstgsssrdsas 623

RESULT 14
AAB54149
ID AAB54149 standard; Protein: 306 AA.

XX AAB54149;
AC
DT 09-MAR-2001 (first entry)
XX

DE Human pancreatic cancer antigen protein sequence SEQ ID NO:601.
XX

KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW detection; diagnosis; identification; cytostatic; neuroprotective;
KW neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative.
XX
OS Homo sapiens.
XX
PN WO200055320-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05989.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI: 2000-579444/54.
XX
DR N-PSDB: AAC98914.
XX

PT New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particular pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition -
PT
XX
XX Claim 11; Page 1037-1038; 1379pp; English.

CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytosolic,
CC neuroprotective, neotropic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiant and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC

CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.

XX Sequence 306 AA;

Query Match 26.6%; Score 67; DB 21; Length 306;
Best Local Similarity 37.0%; Pred. No. 2.4;
Matches 20; Conservative 9; Mismatches 17; Indels 8; Gaps 2;

OY 2 TGHSGSOHSHHTTTCGRSDA-----SRGSSGSRSTSRRTDQDQSGDGR--HS 47
Db 140 sgsskshsksrstsrsgsrstsrsgsrstsrsgsrstsrsgsrstsrsgsrstsrshs 193
:| | | | | : : : | | | | | : : : | | | | |

RESULT 15

AAAM40069
ID AAAM40069 standard; Protein: 443 AA.

XX AC AAAM40069;

XX DT 22-OCT-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 3214.

XX KW Human; nootropic; immunosuppressant; cytoskeletal; gene therapy; cancer;
XX KW peripheral nervous system; neuropathy; central nervous system; CNS;
XX KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX KW leukaemia.

XX OS Homo sapiens.

XX PN WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US34263.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-0552317.

XX PR 09-JUL-2000; 2000US-0598042.

XX PR 19-JUL-2000; 2000US-0620312.

XX PR 03-AUG-2000; 2000US-0653450.

XX PR 14-SEP-2000; 2000US-0662191.

XX PR 19-OCT-2000; 2000US-0693036.

XX PR 29-NOV-2000; 2000US-0727344.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX PI Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX DR WPI: 2001-442253/47.

XX DR N-PSDB; AAI59225.

XX PT Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries -

XX PS Example 5; SEQ ID NO 3214; 10078bp; English.

XX CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
XX CC the encoded polypeptides (AAAM38642-AAAM42213) with nootropic,
XX CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and

CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX SQ Sequence 443 AA;

Query Match 26.6%; Score 67; DB 22; Length 443;
Best Local Similarity 38.9%; Pred. No. 3.6;
Matches 21; Conservative 8; Mismatches 19; Indels 6; Gaps 3;

OY 1 STGHSGSOHSHHTT--QGRSDASRGSSGSRSTSRRTDQDQSG--DGRHSGS 49
Db 377 ntssnrtrsgwsggrsggrsggrsggrsggrs-grgsrgrsgsrgrdgrrrsgn 429
:| : | | : : | | | | | | | : : | | | | |

Search completed: January 28, 2002, 07:39:28
Job time: 118 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: January 28, 2002, 07:37:30 ; Search time 12.62 Seconds
(without alignments)
87.374 Million cell updates/sec

Title: US-09-582-711-3

Perfect score: 252

Sequence: 1 STGSGSGHSHHTTQGRSDA.....TSRETRDQEGSGDSRHSGS 49

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing First 45 summaries

Database :

Issued Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6C.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	73	29.0	743	4	US-08-910-925-3
2	70.5	28.0	1261	4	US-09-208-742-4
3	70.5	28.0	1261	4	US-09-332-295-2
4	69.5	27.6	703	4	US-08-910-925-4
5	69	27.4	717	4	US-08-910-925-1
6	65	25.8	586	3	US-08-964-268-3
7	64	25.4	3025	6	523423-3
8	61.5	24.4	1093	3	US-08-545-860D-55
9	61.5	24.4	1093	5	PCT-US94-04496-55
10	61	24.2	1581	4	US-09-110-517-2
11	60	23.8	984	1	US-08-257-073-3
12	60	23.8	984	2	US-08-184-009-120
13	60	23.8	984	2	US-08-458-356-120
14	60	23.8	984	4	US-08-460-736-120
15	57.5	22.8	514	4	US-08-796-899-25
16	57	22.6	414	1	US-07-667-276A-4
17	56.5	22.4	73	1	US-08-272-882D-3
18	56.5	22.4	73	5	PCT-US92-05532-3
19	56.5	22.4	1481	2	US-08-616-844-40
20	56.5	22.4	1481	2	US-08-599-654-40
21	56.5	22.4	1481	3	US-08-944-868A-40
22	56.5	22.4	1481	3	US-08-944-423A-40
23	56.5	22.4	1481	3	US-08-944-496-40
24	56.5	22.4	3969	4	US-08-061-376-5
25	56	22.2	376	4	US-09-125-099-10
26	56	22.2	376	4	US-09-125-099-12
27	56	22.2	400	5	PCT-US95-16472-2

28	56	22.2	402	3	US-08-602-809-2	Sequence 2, App11
29	56	22.2	568	1	US-08-320-559-30	Sequence 30, App1
30	56	22.2	568	3	US-08-545-860D-30	Sequence 30, App1
31	56	22.2	568	5	PCT-US94-04496-30	Sequence 30, App1
32	55.5	22.0	401	2	US-08-549-004A-5	Sequence 5, App11
33	55.5	22.0	401	4	US-09-051-982A-5	Sequence 5, App11
34	55.5	22.0	410	2	US-08-723-415B-10	Sequence 10, App1
35	55.5	22.0	410	2	US-08-723-415B-11	Sequence 11, App1
36	55.5	22.0	410	2	US-08-428-131-2	Sequence 2, App11
37	55.5	22.0	410	2	US-08-602-846-2	Sequence 2, App11
38	55.5	22.0	410	4	US-09-078-596-2	Sequence 2, App11
39	55.5	22.0	410	4	US-09-189-627A-10	Sequence 10, App1
40	55.5	22.0	410	4	US-09-189-627A-11	Sequence 10, App1
41	55.5	22.0	763	2	US-08-677-862-2	Sequence 2, App11
42	55.5	22.0	763	2	US-09-252-571-2	Sequence 2, App11
43	55.5	22.0	763	3	US-09-434-065-2	Sequence 2, App11
44	55.5	22.0	763	4	US-08-789-275-4	Sequence 4, App11
45	55.5	22.0	763	4	US-08-789-275-5	Sequence 5, App11

ALIGNMENTS

RESULT 1
US-08-910-925-3
; Sequence 3, Application US/08910925
; Patent No. 6162601
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN PININ SPLICE VARIANT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910, 925
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0365 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 743 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1684847
; US-08-910-925-3
Query Match 29.0%; Score 73; DB 4; Length 743;


```

APPLICATION NUMBER: US/08/910,925
FILING DATE: Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PE-0365 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 717 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: FIBROT01
CLONE: 53219
US-08-910-925-1

Query Match          27.4%; Score 69; DB 4; Length 717;
Best Local Similarity 38.8%; Pred. No. 1.1;
Matches 19; Conservative 6; Mismatches 22; Indels 2; Gaps 1.;

QY      1 STGHGSOHSHITTTGGRSDASRGSGSRSTSHETRDQEOCGGSRHSGS 49
       1 : - - - - - : - - - - - : - - - - - : - - - - - :
Db      577 SSSSSSSSSSTSSSSGGSSSSGGSSSSSSSRSSS--SSSSSTGGSSSPDSSS 623

RESULT        6
US-08-964-268-3
Sequence 3, Application US/08964268
Patent No. 6114503
GENERAL INFORMATION:
APPLICANT: WEI, YING-FEI
APPLICANT: RUBEN, STEVEN M
APPLICANT: SANCAR, AZIZ
APPLICANT: HSU, SHIAO-WEN D
APPLICANT: KAZANTSEV, ALEKSEY G
TITLE OF INVENTION: HUMAN BLUE-LIGHT PHOTORECEPTOR hCRY2
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/Ms-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/964,268
FILING DATE: HEREMITH
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/030,189
FILING DATE: 04-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0750001/EKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 3:

```

```

SEQUENCE CHARACTERISTICS:
; LENGTH: 586 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-964-268-3

Query Match      25.8%; Score 65; DB 3; Length 586;
Best Local Similarity 41.2%; Pred. No. 2.7;
Matches 14; Conservative 3; Mismatches 17; Indels 0; Gaps 0;

OY   4 HSGSHSTTTQGRDASRGSSGSRSTRERDQ 37
      |  || : | |||| | : | : | : |
Db    541 HGDSQOTHILKQGRSSMGTGLSGCKRPSQEDDTQ 574

RESULT 7
5223423-3
; Patent No. 5223423
; APPLICANT: FRANCHINI, GENOVEPPA,WONG-STPAAL, FLOSSIE;
; GALLO, ROBERT
; TITLE OF INVENTION: CHARACTERIZATION OF REPLICATION COMPETENT
; HUMAN IMMUNODEFICIENCY TYPE 2 PROVIRAL CLONE HIV-2 SBL/ISY
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/331,212
; FILING DATE: 03-31-1989
; SEQ ID NO.:3:
; LENGTH: 3025
5223423-3

Query Match      25.4%; Score 64; DB 6; Length 3025;
Best Local Similarity 27.0%; Pred. No. 24;
Matches 17; Conservative 7; Mismatches 21; Indels 18; Gaps 1;

OY   5 SCGSHSTTTQGRSDASRGSSGSRSTRERTR-----DOEGSGDGSRH 46
      : | | | : | : | : | : | : | : | : | : | : | : | : |
Db    2644 TSSNSETGGDSNQPTCSMGAGSSKKRSRPNGLDRLLRANGACGGAMDSEGGYSQFH 2703

OY   47 SGS 49
      ||
Db    2704 EGS 2706

RESULT 8
US-08-545-860D-55
; Sequence 55, Application US/08545860D
; Patent No. 6040140
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: CanaanI, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-I Region
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &
; ADDRESSEE: No. 6040140r1s
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,860D

```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2002, 07:37:30 : Search time 13.52 Seconds
(without alignments)
276.076 Million cell updates/sec

Title: US-09-582-711-3

Perfect score: 252
Sequence: 1 STGSGSGSHHTTQGRSDA.....TSRETRDQSGSGSRHSGS 49

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	252	100.0	2248	2 A35938	profilaggrin - hum
2	207	82.1	416	2 A32947	filaggrin precursor
3	207	82.1	591	2 A45135	profilaggrin - hum
4	77	30.6	343	2 T05221	hypothetical prote
5	73.5	29.2	716	2 S43693	pericollin-binding
6	73.5	29.2	1507	2 B47328	natural killer cel
7	72.5	28.8	305	2 TC4525	nucleic acid-bind
8	71	28.2	196	2 T29343	hypothetical prote
9	70.5	28.0	819	2 T08745	probable RNA helic
10	69.5	27.6	471	2 T33997	hypothetical prote
11	69.5	27.6	1217	2 S52714	sericlinb - silkw
12	69	27.4	174	2 S18654	variant surface an
13	69	27.4	284	2 A35419	neutrophil protein
14	69	27.4	625	2 A34615	profilaggrin - rat
15	68.5	27.2	254	2 A31488	filaggrin - mouse
16	68.5	27.0	313	2 A28444	filaggrin precursor
17	68	27.0	406	2 S38170	SRP40 protein - ye
18	68	27.0	1217	2 T24625	AF-4 protein - mou
19	67.5	26.8	156	2 H34768	ORF8 protein - Orf
20	67	26.6	399	2 T46259	hypothetical prote
21	67	26.6	494	2 A48133	pre-mRNA splicing
22	67	26.6	669	2 T44681	GTP-binding protei
23	66.5	26.4	635	2 F75477	hypothetical prote
24	66	26.2	860	2 JC5702	ErkB kinase activa
25	66	26.2	868	2 JC5701	ErkB kinase activa
26	65.5	26.0	115	2 D61615	sericin Mg-1 - gre
27	65.5	26.0	294	2 T35892	hypothetical prote
28	65.5	26.0	1207	2 T23754	hypothetical prote
29	65.5	26.0	1469	2 T09219	basal transcriptio

30	65	25.8	635	2 T00011	ceal protein - rat
31	65	25.8	772	2 T27512	hypothetical prote
32	65	25.8	972	2 S35521	DNA topoisomerase
33	64.5	25.6	190	2 A24713	sericin - silkworm
34	64.5	25.6	492	2 S63065	hypothetical prote
35	64.5	25.6	510	2 T37541	probable glycolipl
36	64	25.4	250	2 B35026	filaggrin B - mous
37	64	25.4	255	2 A35026	filaggrin A - mous
38	64	25.4	292	2 T23966	hypothetical prote
39	64	25.4	344	2 T40167	hypothetical prote
40	64	25.4	505	2 T05502	hypothetical prote
41	64	25.4	764	2 T39878	hypothetical prote
42	64	25.4	1788	2 T31095	vitellogenin I pre
43	64	25.4	1912	2 T29088	vitellogenin I pre
44	63.5	25.2	328	2 T00747	RING-H2 finger pro
45	63.5	25.2	438	2 C86244	DnaH homolog, 4706

ALIGNMENTS

```

RESULT 1
A35938
profilaggrin - human (fragments)
C:Species: Homo sapiens (man)
C>Date: 14-Dec-1990 #sequence_revision 02-Jul-1996 #text_change 29-Sep-1999
C:Accession: A35938
R:Gan, S.O.; McBride, O.W.; Idler, W.W.; Markova, N.; Steinert, P.M.
Biochemistry 29, 9432-9440, 1990
A:Title: Organization, structure, and polymorphisms of the human profilaggrin gene.
A:Reference number: A35938; MUID:91064347
A:Accession: A35938
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-2248 <GAN>
A:Cross-references: GB:J02929
C:Genetics:
A:Gene: GDB:FLG
A:Cross-references: GDB:119912; OMIM:135940
A:Map position: 1q21-1q21
C:Superfamily: unassigned calmodulin-related proteins; calmodulin repeat homology
C:Keywords: EF hand; epidermis; polymorphism; tandem repeat
F:246-569/Region: filaggrin repeat
F:570-893/Region: filaggrin repeat
F:1074-1397/Region: filaggrin repeat
F:1573-1896/Region: filaggrin repeat

Query Match 100.0% Score 252; DB 2; Length 2248;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 STGSGSGSHHTTQGRSDASRGSGSRSTRETRDQSGSGSRHSGS 49
DB 316 STGSGSGSHHTTQGRSDASRGSGSRSTRETRDQSGSGSRHSGS 364

RESULT 2
A32947
filaggrin precursor - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 20-Dec-1988 #sequence_revision 04-Sep-1992 #text_change 29-Sep-1999
C:Accession: A32947
R:McKinnley-Grant, L.J.; Idler, W.W.; Bernstein, I.A.; Parry, D.A.D.; Cannizzaro, L.;
Proc. Natl. Acad. Sci. U.S.A. 86, 4848-4852, 1989
A:Title: Characterization of a cDNA clone encoding human filaggrin and localization o
A:Reference number: A32947; MUID:89296901
A:Accession: A32947
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-416 <MCK>
A:Cross-references: GB:M24355; NID:9182604; PIDN:AA52454.1; PID:9182605
A>Note: the authors translated the codon CAC for residue 188 as Gln, and AAT for res1

```

C:Genetics:
A:Gene: GDB:FLC
A:Cross-references: GDB:119912; OMIM:135940
A:Map position: 1q21-1q21
C:Superfamily: unassigned calmodulin-related proteins; calmodulin repeat homology
C:Keywords: EF hand; epidermis; polymorphism; tandem repeat

Query Match 82.1%; Score 207; DB 2; Length 416;
Best Local Similarity 81.2%; Pred. No. 3.5e-15;
Matches 39; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 2 TGHSGSHSHHTTQGRSDASRGSSGSRSTRERDQSGGSRHSGS 49
Db 97 SGHSGSHSHHTTQGRSDASRGSSGSRSTRERDQSGGSRHSGS 144

RESULT 3

A45135
profilaggrin - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 20-Sep-1999
C:Accession: A45135
R:Presland, R.B.; Haydock, P.V.; Fleckman, P.; Nirunskitiri, W.; Dale, B.A.
J. Biol. Chem. 267, 23772-23781, 1992
A:Title: Characterization of the human epidermal profilaggrin gene. Genomic organization
A:Reference number: A45135; MUID:93054736
A:Accession: A45135
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-591 <PRE>
A:Cross-references: GB:L01089; GB:M90967; NID:g190408; PIDN:AAA60177.1; PID:g553621
A:Note: sequence extracted from NCBI backbone (NCBIP:118773).
C:Genetics:

A:Gene: GDB:FLG
A:Cross-references: GDB:119912; OMIM:135940
A:Map position: 1q21-1q21
C:Superfamily: unassigned calmodulin-related proteins; calmodulin repeat homology
C:Keywords: EF hand; epidermis; polymorphism; tandem repeat
F:49-81/Domain: calmodulin repeat homology <EF2>

Query Match 82.1%; Score 207; DB 2; Length 591;
Best Local Similarity 79.2%; Pred. No. 4.9e-15;
Matches 38; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 2 TGHSGSHSHHTTQGRSDASRGSSGSRSTRERDQSGGSRHSGS 49
Db 539 SGHSGSHSHHTTQGRSDASRGSSGSRSTRERDQSGGSRHSGS 586

RESULT 4

T05221
hypothetical protein F17I5.120 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C:Accession: T05221
R:Bevan, M.; Vitale, D.; Liguori, R.; Arifitou, A.; De Simone, V.; Hohelsel, J.; Mewes,
submitted to the Protein Sequence Database, July 1998
A:Reference number: Z15404
A:Accession: T05221
A:Molecule type: DNA
A:Residues: 1-343 <BEV>
A:Cross-references: EMBL:AL031032
C:Genetics:
A:Experimental source: cultivar Columbia; BAC clone F17I5
A:Map position: 4
A:Note: F17I5.120

Query Match 30.6%; Score 77; DB 2; Length 343;
Best Local Similarity 42.0%; Pred. No. 0.3;
Matches 21; Conservative 4; Mismatches 19; Indels 6; Gaps 2;

OY 5 SSGSHSH---TTTQGRSDASRGSSGSRSTR---ETRDQSGGDSRHSG 48
Db 89 SSGSHSHSGTTHNGSHSSGSHSSATGSTRHNGHTSTGSHSSGSHSSGSHSS 138

RESULT 5

S43693
penicillin-binding protein 2 - Staphylococcus aureus
C:Species: Staphylococcus aureus
C:Date: 10-Dec-1994 #sequence_revision 10-Nov-1995 #text_change 18-Feb-2000
C:Accession: S43693
R:Murakami, K.; Fujimura, T.; Doi, M.
FEBS Microbiol. Lett. 117, 131-136, 1994
A:Title: Nucleotide sequence of the structural gene for the penicillin-binding protein
A:Reference number: S43693; MUID:94237445
A:Accession: S43693
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-716 <MUR>
A:Cross-references: EMBL:X62288; NID:g483533; PIDN:CAA44177.1; PID:g483534
C:Superfamily: penicillin-binding protein 1B

Query Match 29.2%; Score 73.5; DB 2; Length 716;
Best Local Similarity 45.9%; Pred. No. 1.5;
Matches 17; Conservative 6; Mismatches 11; Indels 3; Gaps 1;

OY 5 SSGSHSH---TTTQGRSDASRGSSGSRSTRERDQSGGSRHSGS 38
Db 679 SSGSHSHSHHTTQGRSDASRGSSGSRSTRERDQSGGSRHSGS 715

RESULT 6

B47328
natural killer cell tumor-recognition protein - mouse
N:Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-t
C:Species: Mus musculus (house mouse)
C:Date: 17-Jul-1994 #sequence_revision 17-Jul-1994 #text_change 05-Nov-1999
C:Accession: B47328; I77662
R:Anderson, S.K.; Gallinger, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R.
Proc. Natl. Acad. Sci. U.S.A. 90, 542-546, 1993
A:Title: A cyclophilin-related protein involved in the function of natural killer cell
A:Reference number: A47328; MUID:93133824
A:Accession: B47328
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1507 <AND>
A:Cross-references: GB:L04289; NID:g192866
A:Note: authors translated the codon AGT for residue 972 as Arg
R:Rinfret, A.; Anderson, S.K.
Mol. Immunol. 30, 1307-1313, 1993
A:Title: IL-2 regulates the expression of the NK-TR gene via an alternate RNA splicing
A:Reference number: I57820; MUID:94019422
A:Accession: I77662
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 235-237/263-294 <RTN>
A:Cross-references: GB:S65998; NID:g425701; PIDN:AAB28500.1; PID:g425702
C:Genetics:
A:Gene: NK-TR
C:Superfamily: natural killer cell tumor-recognition protein; cyclophilin homology
C:Keywords: alternative splicing; lymphocyte
F:60-230/Domain: cyclophilin homology <CYF>

Query Match 29.2%; Score 73.5; DB 2; Length 1507;
Best Local Similarity 39.3%; Pred. No. 3.1;
Matches 22; Conservative 6; Mismatches 21; Indels 7; Gaps 2;

OY 1 STGHSGSHSHHTTQGRS---DASRGSSGSRSTRERDQSGGDSRHSGS 49
Db 740 SSGSGSKSSSRSSSRSTRSRSLPTSRSLSPSSRSHSPMKYSDGSHSGS 795

RESULT 7
 Jc4525
 nucleic acid-binding protein E5.1 - human
 N:Alternate names: E5.1 protein
 C:Species: Homo sapiens (man)
 C:Date: 08-Feb-1996 #sequence_revision 19-Apr-1996 #text_change 24-Sep-1999
 C:Accession: Jc4525
 R:Badolato, J.; Gardiner, E.; Morrison, N.; Eisman, J.
 Gene 166, 323-327, 1995
 A:Title: Identification and characterization of a novel human RNA-binding protein.
 A:Reference number: Jc4525; MUID:96125212
 A:Accession: Jc4525
 A:Molecule type: mRNA
 A:Residues: 1-305 <BAD>
 A:Cross-references: GB:L37366; NID:g1236282; PIDN:AAA92859.1; PID:g598231
 C:Comment: This protein functions in RNA metabolism, and is involved in processing of pre-mRNA.
 C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein repeat-containing proteins
 C:Keywords: RNA binding
 F:69-121/Region: serine-rich
 F:128-154,269-298/Region: arginine-rich
 F:162-241/Domain: RNA binding #status predicted <RNM>
 F:162-230/Domain: ribonucleoprotein repeat homology <RRM1>
 F:163-168/Region: ribonucleoprotein-2 heptamer
 F:203-210/Region: ribonucleoprotein-1 octamer
 F:278-287/Region: nuclear location signal

Query Match 28.8%; Score 72.5; DB 2; Length 305;
 Best Local Similarity 43.1%; Pred. No. 0.82;
 Matches 22; Conservative 4; Mismatches 20; Indels 5; Gaps 2;

QY 1 STGHSSQSHHTTQGRSDASRGS-SGSRSTSRRTDQSGDGS---RH 46
 DB 81 STSSGSSSTSTSSGSSSSSSASRSGSSSTSRSSSSSSSSGSPSPRRRH 131

RESULT 8
 T29343
 hypothetical protein F21C10.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T29343
 R:Du, Z.; Gattung, S.
 Submitted to the EMBL Data Library, April 1996
 A:Description: The sequence of C. elegans cosmid F21C10.
 A:Reference number: Z20610
 A:Accession: T29343
 A:Molecule type: DNA
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Residues: 1-196 <DUZ>
 A:Cross-references: EMBL:U55364; PIDN:AAA97976.1; GSPDB:GN00023; CESP:F21C10.4
 A:Experimental source: strain Bristol N2; clone F21C10
 C:Genetics:
 A:Gene: CESP:F21C10.4
 A:Map position: 5
 A:Introns: 59/2; 106/2; 178/2

Query Match 28.2%; Score 71; DB 2; Length 196;
 Best Local Similarity 47.1%; Pred. No. 0.77;
 Matches 16; Conservative 3; Mismatches 15; Indels 0; Gaps 0;

QY 16 GRSDASRSGSGSRSTSRRTDQSGDGSRHSGS 49
 DB 27 GRGGGRCGSSGARSSRSRSGAGAGKFRHSSS 60

RESULT 9
 T08745
 probable RNA helicase - human
 N:Alternate names: protein DKF2p586A0419.1

C:Species: Homo sapiens (man)
 C:Date: 11-Jun-1999 #sequence_revision 18-Aug-2000 #text_change 19-Jan-2001
 C:Accession: T08745; T13148
 R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, March 1999
 A:Reference number: T16471
 A:Accession: T08745
 A:Molecule type: mRNA
 A:Residues: 455-819 <WAM>
 A:Cross-references: EMBL:AL050096; PIDN:CAB43268.1
 A:Experimental source: adult uterus; clone DKF2p586A0419
 R:Suk, K.; Lee, M.S.
 submitted to the EMBL Data Library, August 1998
 A:Description: A novel human cDNA homologous to RNA helicase genes.
 A:Reference number: Z17605
 A:Accession: T13148
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-698; 'PSOSRRDQMSA' <SUK>
 A:Cross-references: EMBL:AF083255; NID:g3435311; PID:g3435312; PIDN:AMC32396.1
 C:Superfamily: ATP-dependent RNA helicase DBP1
 C:Keywords: ATP; DEAD box; nucleotide binding; P-loop
 F:178-185/Region: nucleotide-binding motif A (P-loop)
 F:284-289/Region: nucleotide-binding motif B
 F:288-291/Region: DEAD motif

Query Match 28.0%; Score 70.5; DB 2; Length 819;
 Best Local Similarity 34.1%; Pred. No. 3.6;
 Matches 15; Conservative 8; Mismatches 16; Indels 5; Gaps 1;

QY 3 GHSGSQSHHTTQGRSDASRSGSGSRSTSRRTDQSGDGSRH 46
 DB 683 GSNKRRERYTENKSSSRHSHGETGNRHS-----DSPRHGGGRH 721

RESULT 10
 T33997
 hypothetical protein W03G1.5 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T33997
 R:Pauley, A.; Scheel, P.; Harper, M.
 submitted to the EMBL Data Library, February 1999
 A:Description: The sequence of C. elegans cosmid W03G1.
 A:Reference number: Z21454
 A:Accession: T33997
 A:Molecule type: DNA
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Residues: 1-471 <PAU>
 A:Cross-references: EMBL:AF125964; PIDN:AD14753.1; GSPDB:GN00022; CESP:W03G1.5
 A:Experimental source: strain Bristol N2; clone W03G1
 C:Genetics:
 A:Gene: CESP:W03G1.5
 A:Map position: 4

Query Match 27.6%; Score 69.5; DB 2; Length 471;
 Best Local Similarity 40.4%; Pred. No. 2.7;
 Matches 21; Conservative 7; Mismatches 17; Indels 7; Gaps 2;

QY 1 STGHSSQSHHTTQGRSDASRSGSGSRSTSRRTDQSGDGSRHSGS 49
 DB 120 SSSSFSKVSF---GRSESSSSSGESSDDENRRKKKGRSGSSSSSGS 167

RESULT 11
 S52714
 sericin1B - silkworm
 C:Species: Bombyx mori (silkworm)
 C:Date: 19-May-1995 #sequence_revision 21-Jul-1995 #text_change 09-Sep-1997
 C:Accession: S52714
 R:Garel, A.A.; Deleage, G.G.; Prudhomme, J.J.

submitted to the EMBL Data Library, March 1995
A:Description: Structure and organisation of the Bombyx mori sericin I gene and of the f
A:Reference number: S52714
A:Accession: S52714
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1217 <GAR>
A:Cross-references: EMBL:248802; NID:9755699; PTD:9755700

Query Match	27.6%	Score 69.5	DB 2	Length 1217
Best Local Similarity	40.8%	Pred. No. 6.9		
Matches	20	Conservative	6	Mismatches 22; Indels 1; Gaps 1;

QY	1	STGHSQSHS--HTTQGRSDASRSGSGSRGSRYSRETRDQESDGSGRHSG	48
Db	749	STKNGSTSGCSTIYGRSSSRHGGSVSTGSSNTDSTKNAGRTSG	797

RESULT 12

S18654

variant surface antigen vlpB precursor - Mycoplasma hyorhinis

N:Alternate names: lipoprotein

C:Species: Mycoplasma hyorhinis

C:Date: 22-Nov-1993 #sequence_revision 14-Jul-1995 #text_change 21-Jul-2000

C:Accession: S18654

R:Yogev, D.; Rosengarten, R.; Watson-McKown, R.; Wise, K.S.

EMBO J. 10, 4069-4079, 1991

A:Title: Molecular basis of Mycoplasma surface antigenic variation: a novel set of diver

A:Reference number: S18651; MUID:92097525

A:Accession: S18654

A:Molecule type: DNA

A:Residues: 1-174 <YOG>

A:Cross-references: EMBL:X62936; NID:94495092; PIDN:CAA44709.1; PID:g581323

C:Genetics:

A:Gene: vlpB

A:Genetic code: SGC3

A:Start codon: GTG

C:Superfamily: glycine-rich cell wall structural protein 1

C:Keywords: surface antigen

F:1-29/Domain: signal sequence #status predicted <SIG>

F:30-174/Product: variant surface protein vlpB #status predicted <MAT>

```

Query Match 27.4%: Score 69; DB 2; Length 174;
Best Local Similarity 43.1%: Pred. NO. 1.1;
Matches 22; Conservative 7; Mismatches 20; Indels 2; Gaps 2.

OY 1 STGAGS-QHSHTTQGRSDASRGS-SCSRSTRRETRPOEGSGDGRHSGS 49
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48 SGGGSGTLAGSGTTTGGSGTTTGGRRSSSSSSSTTGGGQTGTGSDSDSGA 98
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 13
A35419
neutrophil protein - pig (fragment)
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 18-Jan-1991 #sequence_revision 18-Jan-1991 #text_change 05-Nov-1999
C:Accession: A35419
R:Belaville, P.; Bazzoni, F.; Cassatella, M.A.; Hunter, K.J.; Bannister, J.V.
Biochem. Biophys. Res. Commun. 170, 915-922, 1990
A:Title: Isolation and characterization of a cDNA clone for a novel serine-rich neutrophil
A:Reference number: A35419; MUID:90343818
A:Accession: A35419
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-284 <BEL>
A:Cross-references: GB:M55701; NID:g164672; PIDN:AAA63449.1; PID:g164673

```

Query Match	27.48;	Score 69;	DB 2;	Length 284;
Best Local Similarity	43.28;	Pred. NO. 1.8;		
Matches	16;	Conservative	6;	Mismatches 15;
				Indels 0;
				Gaps 0;

QY 1 STGHCQSQHSHHTTQGRSDASRGSSSSRSTSTRTDQ 37
| : | | | : : | | | | | : | |
Db 144 SSSSSSSSSSTSSSGSSSSSSGSSSSRSSSSSSSDQ 180

RESULT 14
A34615
profilaggrin - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 29-Jun-1990 #sequence_revision 03-Oct-1992 #text_change 04-Feb-2000
C:Accession: A34615
R:Haydock, P. V.; Dale, B.A.
DNA Cell Biol. 9, 251-261, 1990
A:Title: Filaggrin, an intermediate filament-associated protein: structural and function
A:Reference number: A34615; MUID:50274870

A;Residues: 1-625 <MAY>
A;Cross-references: GB:M21759, MID:9204143, PIDN:AAA41161.1, PID:9204144
A;Note: the authors translated the codon GAA for residue 568 as Gln
C;Superfamily: unassigned calmodulin-related proteins; calmodulin repeat homology
C;keywords: EF hand; epidermis

Query	March	Similarity	28.6%	Score	69	DB	2	Length	625
Best	Local	Similarity	28.6%	Pred.	No.	4			
Matches	11	Conservative	13	Mismatches	22	Indels	0	Gaps	0

OY	1	STGSGS	SH	TT	GT	GR	SD	AS	RG	SS	GS	ST	ST	ET	RD	QD	QD	SG	SR	SG	49
Db	381	ST	RG	SS	QD	AS	ST	AE	IA	RG	GA	VS	EC	PG	FO	RE	AR	SS	HA	OS	429

RESULT 15
A31488
filaggrin - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Jul-1989 #sequence_revision 20-Jul-1989 #text_change 26-May-2000
C:Accession: A31488
R:Resing, K.A.; Walsh, K.A.; Haugen-Scotfield, J.; Dale, B.A.
J. Biol. Chem. 264, 1837-1845, 1989
A:Title: Identification of proteolytic cleavage sites in the conversion of profilaggrin to profilaggrin
A:Reference number: A31488; MUID:89109207
A:Accession: A31488
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-254 <RSS>
C:Superfamily: unassigned calmodulin-related proteins; calmodulin repeat homology
;keywords: EF hand; epidermis

Query Match	27.2%	Score 68.5	DB 2	Length 254
Best Local Similarity	39.6%	Pred. No. 1.9		
Matches 19	Conservative	6	Mismatches 14	Indels 9
Gaps				3
QY	3	GHSSQSHNTTQGRSDA-SKSSSSRSRTSRP----	TRDQEDSGGGR	45
Db	58	GHQANINE-----QGRDSARPRESTNGSHSSRQGGGLSDQDRRSGGGR	101	

Search completed: January 28, 2002, 07:38:59
Job time: 89 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2002, 07:37:30 ; Search time 10.18 Seconds

(without alignments)
176,481 Million cell updates/sec

Title: US-09-582-711-3
Perfect score: 252
Sequence: 1 STGHSGSHSHHTTQGRSDA.....TSRFRDQSGSGSRHSGS 49

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	207	82.1	416	1	FLTA_HUMAN
2	80	31.7	400	1	RTOA_DICDI
3	73.5	29.2	1453	1	NKCR_MOUSE
4	69	27.4	174	1	VLBP_MYCHR
5	68	27.0	406	1	SR40_YEAST
6	68	27.0	1217	1	AF4_MOUSE
7	67	26.6	494	1	SERA_HUMAN
8	67	26.6	510	1	VLPI_BOVIN
9	66.5	26.4	168	1	VLPI_MYCHR
10	66	26.2	868	1	NRG2_RAT
11	65.5	26.0	115	1	SER1_GALME
12	65	25.8	972	1	TOP1_DROME
13	64.5	25.6	492	1	YNM4_YEAST
14	64.5	25.6	510	1	YDYL_SCHPO
15	64	25.4	336	1	FLTA_MOUSE
16	64	25.4	1912	1	VITL_CHICK
17	63	25.0	486	1	CDSN_HUMAN
18	63	25.0	546	1	PHR2_CANAL
19	63	25.0	817	1	HUNB_MUSDO
20	63	25.0	1127	1	TF1G_HUMAN
21	63	25.0	1850	1	VIT2_CHICK
22	62.5	24.8	326	1	2265_MOUSE
23	62.5	24.8	332	1	2265_MOUSE
24	62.5	24.8	712	1	GSPD_ERWCH
25	62.5	24.8	3726	1	TRX_DROME
26	62	24.6	337	1	2265_HUMAN
27	62	24.6	736	1	CHE1_HUMAN
28	62	24.6	736	1	DVL2_HUMAN
29	62	24.6	851	1	DD21_MOUSE
30	61.5	24.4	757	1	RTS1_YEAST
31	61.5	24.4	989	1	SERA_PLAFG
32	61.5	24.4	1093	1	AF17_HUMAN
33	61.5	24.4	1200	1	DDX8_CAEEL

34	61	24.2	220	1	SER2_GALME	096615 gallieria me
35	61	24.2	549	1	EPD2_CANNA	074137 candida mal
36	61	24.2	588	1	CAR7_CANAL	P43096 candida alb
37	61	24.2	727	1	BRC1_DROME	001295 drosophila
38	61	24.2	736	1	DVL2_MOUSE	060838 mus musculu
39	61	24.2	736	1	DVL2_XENLA	P51142 xenopus lae
40	61	24.2	880	1	BRC4_DROME	024206 drosophila
41	61	24.2	1581	1	PERB_HUMAN	015648 h peroxisom
42	61	24.2	1704	1	VIT1_FUNHE	090508 fundulus he
43	60.5	24.0	345	1	EBR1_RAT	P52796 rattus norv
44	60.5	24.0	389	1	SER1_BOWMO	P07856 bombyx mori
45	60.5	24.0	454	1	VE2_HPV37	080903 human pap11

ALIGNMENTS

RESULT 1
FLTA_HUMAN STANDARD: PRT: 416 AA.

AC P20930: 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE FLAAGRIN PRECURSOR (FRAGMENT).
GN FIG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.

RA MEDLINE-89296901; PubMed-2740331.
RA McInlley-Grant L.J., Idler W.W., Bernstein I.A., Parry D.A.D.,
RA Canizario L., Croce C.M., Huebner K., Lessin S.R., Steinert P.M.;
RT "Characterization of a cDNA clone encoding human filaggrin and
RT localization of the gene to chromosome region 1q21.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:4848-4852(1989)

CC -!- FUNCTION: AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES
CC DISULFIDE-BOND FORMATION AMONG THE INTERMEDIATE FILAMENTS DURING
CC TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.
CC -!- PTM: FILAGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE,
CC HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM COPIES
CC OF 324 AA, WHICH ARE NOT SEPARATED BY "LARGE LINKER". THE
CC PRECURSOR IS DEPOSITED AS KERATOHYALIN GRANULES. DURING TERMINAL
CC DIFFERENTIATION IT IS DEPHOSPHORYLATED & PROTEOLYTICALLY CLEAVED.

CC -----
CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC EMBL: M24355; AAA52454.1; -.
DR PIR: A32947; A32947.
DR MIM: 135940; -.
DR InterPro: IPR003303; Filaggrin.
KW Phosphorylation; Polyprotein; Developmental protein.
FT NON_TER 1
SQ SEQUENCE 416 AA: 44105 MW: DEFA3218BA0A3F32 CMC64;

Query Match 82.1%; Score 207; DB 1; Length 416;
Best Local Similarity 81.2%; Pred. No. 1,86-15;
Matches 39; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 TGHSSGSHSHHTTQGRSDASRGSSRSTRTRDQSGSGSRHSGS 49
DB 97 SGHSGSHSHHTTQGRSDASRGSSRSTRTRDQSGSGSRHSGS 144

[illegible]

```

DE DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DT NK-TUMOR RECOGNITION PROTEIN (NATURAL-KILLER CELLS CYCLOPHILIN-
DE DE RELATED PROTEIN) (NK-TR PROTEIN).
DE OS Mus musculus (Mouse).
DE CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DE CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
DE CC NCBI_TaxID=10090;
DE RN [1]
DE RP SEQUENCE FROM N.A.
DE RX MEDLINE=9313824; PubMed=8421688;
DE RA Anderson S.K., Callinger S., Roder J., Frey J., Young H.A.,
DE RA Ortolano J.R.;
DE RT "A cyclophilin-related protein involved in the function of natural
DE RT killer cells.";
DE RL proc. Natl. Acad. Sci. U.S.A. 90:542-546(1993).
DE RN [2]
DE RP REVISIONS TO C-TERMINUS.
DE RC STRAIN-BALB/C; TISSUE=Blood;
DE RA Anderson S.K.;
DE RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DE CC -1- FUNCTION: COMPONENT OF A PUTATIVE TUMOR-RECOGNITION COMPLEX.
DE CC INVOLVED IN THE FUNCTION OF NK CELLS.
DE CC -1- SIMILARITY: CONTAINS A CYCLOPHILIN-LIKE PRIASE DOMAIN.
DE CC -----
DE CC This SWISS-PROT entry is copyright. It is produced through a collaboration
DE CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
DE CC the European Bioinformatics Institute. There are no restrictions on its
DE CC use by non-profit institutions as long as its content is in no way
DE CC modified and this statement is not removed. Usage by and for commercial
DE CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
DE CC or send an email to license@sib-sib.ch).
DE CC -----
DE CC EMBL; L04289; AAA37500.2; ALT_INIT.
DE DR HSSP; Q27450; 1A58.
DE DR MGD; MGI:97346; Nktr.
DE DR InterPro: IPR002130; CSA_Priase.
DE DR Pfam: PF00160; pro_isomerase; 2.
DE DR PROSITE; PS00170; CSA_PRIASE_1.
DE DR PROSITE; PS50072; CSA_PRIASE_2; 1.
DE RV Cyclosporin; Isomerase; Rotamase; Repeat; Transmembrane.
DE FT DOMAIN 1 176
DE FT DOMAIN 222 241 ARG/LYS-RICH (BASIC).
DE FT DOMAIN 422 459 ARG/LYS-RICH (BASIC).
DE FT DOMAIN 964 1003 ARG/LYS-RICH (BASIC).
DE FT DOMAIN 198 273 ARG/SER-RICH.
DE FT DOMAIN 468 565 ARG/SER-RICH.
DE FT DOMAIN 658 812 ARG/SER-RICH.
DE FT DOMAIN 1303 1453 ARG-SER TANDEM REPEAT-RICH.
DE SO SEQUENCE 1453 AA; 163439 MW; DF1173FF814B283E CRC64;

Query Match 29.2%; Score 73.5; DB 1; Length 1453;
Best Local Similarity 39.3%; Pred. No. 1.6;
Matches 22; Conservative 6; Mismatches 21; Indels 7; Gaps 2;

OY 1 STGHSQSHSTTTQGRS--DASRSGSSGSRSTHETRDQDS---GDGSRHSGS 49
OY I: I I I I I I I I I I I I I I I I I I I I I I I I I I I I
DB 686 SSGRSGSKSSRSRSSRSRYTRSRSLPTSLRSLSPSSRSHSPKMYSDGSGHSRS 741

RESULT 4
ID VLPB_MYCHR STANDARD; PRT; 174 AA.
AC P29229;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE VARIANT SURFACE ANTIGEN B PRECURSOR (VLPB PROLIPROTEIN).
GN VLPB.
OS Mycoplasma hyorhinus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.

```

```

OX NCBI_TaxID=2100;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SK76;
RA MEDLINE=92097525; PubMed=1722868;
RA Yoev D., Rosengarten R., Watson-Mckown R., Wise K.S.;
RT "Molecular basis of Mycoplasma surface antigenic variation: a novel
RT set of divergent genes undergo spontaneous mutation of periodic
RT coding regions and 5' regulatory sequences."
RL EMO J. 10:4069-4079(1991).
CC -1- FUNCTION: RESPONSIBLE FOR THE ANTIGENIC DIVERSITY FOR HOST
CC ADAPTATION.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
CC ANCHOR (PROBABLE).
CC -1- MISCELLANEOUS: THE NUMBERS OF REPEATS CAN VARY AND IS ONE OF THE
CC BASIS OF THE ANTIGENIC DIVERSITY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X62936; CAA44709.1; ALT_SEQ.
DR PIR: S18654; S18654.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN.1.
KW Antigen; Membrane; Lipoprotein; Repeat; Signal.
FT SIGNAL 1 29 PROBABLE.
FT CHAIN 30 174 VARIANT SURFACE ANTIGEN B.
FT LIPID 88 30 7 X 12 AA TANDEM REPEATS.
FT DOMAIN 88 171
FT REPEAT 88 99 1.
FT REPEAT 100 111 2.
FT REPEAT 112 123 3.
FT REPEAT 124 135 4.
FT REPEAT 136 147 5.
FT REPEAT 148 159 6.
FT REPEAT 160 171 7.
SQ SEQUENCE 174 AA; 16145 MW; 28F4C9ECA85585D7 CRC64;

Query Match 27.4%; Score 69; DB 1; Length 174;
Best Local Similarity 43.1%; Pred. NO. 0.5;
Matches 22; Conservative 7; Mismatches 20; Indels 2; Gaps 2;

```

```

RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UBI2 and MPL1 genes and three
RT new open reading frames."
RL Yeast 9:1349-1354(1993).
CC -1- FUNCTION: NOT KNOWN. WEAK SUPPRESSOR OF A MUTANT OF THE
CC SUBUNIT AC40 OF DNA DEPENDANT RNA POLYMERASE I AND III.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L11275; AAA35091.1; -.
DR EMBL: X73541; CAA51946.1; -.
DR EMBL: Z28317; CAA82171.1; -.
DR PIR: S38170; S38170.
DR SCD: S0001800; SRP40.
FT DOMAIN 25 314 ASP/SER-RICH.
FT CONFLICT 400 400 G -> N (IN REF. 1).
SQ SEQUENCE 406 AA; 41015 MW; 8EA007695AF4BAID CRC64;

Query Match 27.0%; Score 68; DB 1; Length 406;
Best Local Similarity 32.7%; Pred. NO. 1.6;
Matches 16; Conservative 8; Mismatches 25; Indels 0; Gaps 0;

OY 1 STGHGSGHSHHTTGGRSASRGSGSRSTRTDQEGSGSGSRHSGS 49
Db 158 SSGSSSSSESGSSESDSDSSSSSSSSSDSDSDSSSSSSSSSDS 206

RESULT 6
AF4_MOUSE STANDARD: PRT: 1217 AA.
AC 088573;
DR 20-AUG-2001 (Rel. 40, Created)
DR 20-AUG-2001 (Rel. 40, Last sequence update)
DR 20-AUG-2001 (Rel. 40, Last annotation update)
DE AF-4 PROTEIN (PROTO-ONCOGENE AF4).
GN MLT12 OR MLT12H.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thymus;
RA Isnard P.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE AF4 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF074266; AAD08668.1; -.
DR MGD: MGI:1100819; MLT12h.
KW Nuclear protein; Proto-oncogene.
FT DOMAIN 261 266 POLY-PRO.
FT DOMAIN 467 476 POLY-SER.
FT DOMAIN 477 483 POLY-GLU.
FT DOMAIN 841 845 POLY-SER.
FT DOMAIN 871 881 POLY-SER.

```


[illegible]

CC	NCBI_Taxid=7137;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Middle silk gland;
RA	Zurrovic M., Sehnaï F., Scheller K., Kumaran A.K.;
RT	"Silk gland specific cDNAs from Galleria mellonella L.";
RL	Insect Biochem. Mol. Biol. 22:55-67(1992).
CC	-1- FUNCTION: PROVIDES THE SILK FIBROIN THREAD WITH A STICKY COATING.
CC	ACTS AS A CEMENT BY STICKING SILK THREADS TOGETHER.
CC	-1- SUBCELLULAR LOCATION: SECRETED.
CC	-1- TISSUE SPECIFICITY: PRODUCED EXCLUSIVELY IN THE MIDDLE (MSG)
CC	SECTION OF SILK GLANDS.
CC	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isdb.ch/announce/
CC	or send an email to license@isb.ch).
CC	
CC	
DR	EMBL; AF095241; AAC79078.1; -
DR	InterPro; IPR000087; Collagen.
DR	Pfam; PF01391; Collagen; 1.
KW	SILK.
FT	NON_TER
SO	SEQUENCE 115 AA; 9101 MW; F6C78FCCA95E4999 CR664;

```

OY      1  STGSGSHSHHTTQGRDASRGSSGSGSSTRETRPDQSGDGRHSS 49
Db      53  SSSSGS--SGSSSGSSGSSGSSGSSGSSGSSSSSSSSSGS--SSGS 98

RESULT 12
TOP1_DROME
AC      ID      TOP1_DROME      STANDARD;      PRT;      972 AA.
AC      P301B9; 09VXW6;
DT      01-APR-1993 (Rel. 25, Created)
DT      01-APR-1993 (Rel. 25, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      DNA TOPOISOMERASE I (Ec 5.99.1.2).
DE      TOP1 OR CG6146.
OS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC      Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC      Ephydroidea; Drosophilidae; Drosophila.
OX      NCBI_TaxID=7227;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=93117086; PubMed=1335568;
RA      Hsieh T.-S., Brown S.D., Huang P., Fostel J.;
RT      "Isolation and characterization of a gene encoding DNA topoisomerase
RT      I in Drosophila melanogaster."
RL      Nucleic Acids Res. 20:6177-6182(1992).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=96354910; PubMed=8769417;
RA      Zhang C.X., Lee M.P., Chen A.D., Brown S.D., Hsieh T.-S.;
RT      "Isolation and characterization of a Drosophila gene essential for
RT      early embryonic development and formation of cortical cleavage
RT      furrows."
RL      J. Cell Biol. 134:923-934(1996).
RN      [3]
RP      SEQUENCE FROM N.A.
RX      STRAIN=BERKELEY;
RA      MEDLINE=20196006; PubMed=10731132;
RA      Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

```

RA Amandides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwatz R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bandaru D., Bolshakov S.,
RA Butkova D., Butchan M.R., Bouck J., Brockstein P., Brotler P.,
RA Borika C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garq N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibbegam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kratiz C., Kravitz S., Kulp D., Lal Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattet B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekullov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murry L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sliden-Klamas I., Simpson M., Skupski M.P., Smith T.,
RA Splier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svyrkas R., Teector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Wolley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhao M., Zhang G., Zhao Q., Zheng L.,
RA Zhong X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
CC
CC -I- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.
CC
CC -I- CATALYTIC ACTIVITY: ATP-INDEPENDENT BREAKAGE OF SINGLE-STRANDED
CC DNA, FOLLOWED BY PASSAGE AND REJOINING.
CC
CC -I- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC RELAX ONLY NEGATIVE SUPERCOILS.
CC
CC -I- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN
CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.
CC
CC -I- SIMILARITY: BELONGS TO THE EUKARYOTIC TYPE I TOPOISOMERASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U74557; AAA28951.1; -.
DR EMBL; M80064; AAC24158.1; -.
DR EMBL; AE003498; AAP48440.1; -.
DR PIR; S35521; S35521.
DR HSSP; P11387; 1A35.
DR FlyBase; FBgn0004924; Top1.
DR InterPro; IPR001631; Topoisomerase_I.
DR Pfam; PF01028; Topoisomerase_I.
DR PRINTS; PR00416; EUTPISMRASET.
DR SMART; SM00435; TOPEUC; 1.
DR PROSITE; PS00176; TOPOISOMERASE_I_EUK; 1.
KW Isomerase; Topoisomerase; DNA-binding.
FT DOMAIN 32 39 POLY-HIS.
FT DOMAIN 40 198 SER-HIS.
FT ACT_SITE 930 930 DNA CLEAVAGE (BY SIMILARITY).

FT	CONFLICT	40	40	S -> H (IN REF. 3).
FT	CONFLICT	46	46	S -> SSS (IN REF. 3).
FT	CONFLICT	201	201	H -> O (IN REF. 3).
SQ	SEQUENCE	972 AA;	11688 MW;	3764BBBDEDEA30CD CRC64

Query Match	25.8%	Score	65;	DB	1;	Length	972;
Best Local Similarity	35.1%	Pred. No.	84;				
Matches	20;	Conservative	5;	Mismatches	24;	Indels	8;
						Gaps	3;

```

QY      1  STGH-SGSDSHSTTTTQGRSDASGSSGSR----STSPETRDQEQ---SGDGRHS 49
          | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     105  SNSHRSGSSSHKDXDGGSSSSSKHKSSSGHHKRSSKDXERDCKDXDRCSSSSSRHKSS 161

```

```

RESULT 13
YNM4_YEAST
ID YNM4_YEAST STANDARD; PRT; 492 AA.

```

DT 01-OCT-1996 (Rel. 34, last sequence update)
DT 15-JUL-1998 (Rel. 36, last annotation update)
DE HYPOTHETICAL 54.9 KDA PROTEIN IN SPC98-TOM70 INTERGENIC REGION.

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;

RX MEDLINE-97245296; PubMed-9090055;
RA de Antoni A., D Angelo M., Dal Pero F., Sartorello F., Pandolfo D.
RA Pallevicini A., Lanfranchi G., Valle G.;

RT overlapping open reading frames.";
RL Yeast 13:261-266(1997).
RN [2]

RX MEDLINE=96109932; PubMed=8619318;
RA Mallet L., Bussereau F., Jacquet M.;
RT "A 43.5 kb segment of yeast chromosome XIV, which contains MFA2,

```

RL  Yeast 11:1195-1209(1995).
CC  -I- SIMILARITY: TO S.POMBE SPBC30D10.15
CC  -----

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC	EMBL: 269382; CAA93381.1; -
DR	EMBL: 271400; CAA96005.1; -

SEQUENCE 492 AA; 54949 MW; 5A496278070F6255 CRC64

QY 1 SYG-HSGSQSHSTTTGCRSDAAGSGSGSR--STSRRTDDEQSGD 42
Matches 17; Conservative 8; Mismatches 17; Indels 3; Gaps 2;

RESULT 14

```

YVDL_SCHPO
ID YVDL_SCHPO STANDARD; PRT; 510 AA.
AC 013692;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOHECTICAL 53.7 KDA PROTEIN C11E3.13C IN CHROMOSOME I PRECURSOR.
GN SPAC11E3.13C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
NCBI_TaxID=4896;
OX
RN
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RL Murphy L., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
-1- SIMILARITY: BELONGS TO THE GAS1 FAMILY.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC CC EMBL: Z98595; CAB1192.1; -
CC
DR KW Hypothetical protein; Glycoprotein; Signal.
FT CHAIN 1 22
FT SIGNAL 1 22
FT DOMAIN 423 510
FT CARBOHYD 147 147 SER-RICH.
FT CARBOHYD 216 216 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 337 337 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 397 397 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 491 491 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 510 AA; 53692 MW; 3D515CCDB5651087 CRC64;

Query Match 25.6%; Score 64.5; DB 1; Length 510;
Best Local Similarity 38.0%; Pred. No. 4.8;
Matches 19; Conservative 5; Mismatches 25; Indels 1; Gaps 1;

OY 1 STGSSGSHHTTQGRSDASRGSS-GSRSTRERDQEGSGDSRRHSGS 49
DB 427 TSGSSGSSASSTTASSSVSSGSSISSGSSMSSTYTSASGSSAHSSGS 476

RESULT 15
FILE_MOUSE STANDARD; PRT; 336 AA.
AC P11088;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE FLAGGRIN (FRAGMENT).
GN FLAG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
OX
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=88058903; PubMed=3680218;
RA Rothnagel J.A., Mehrel T., Idler W.W., Roop D.R., Steinert P.M.;
RT "The gene for mouse epidermal flaggrin precursor. Its partial
RT characterization, expression, and sequence of a repeating flaggrin
RT unit."
RL J. Biol. Chem. 262:15643-15648(1987).

```

```

RN [2]
RP REVISIONS.
RA Rothnagel J.A.;
RL Submitted (SEP-1988) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES
CC DISULFIDE-BOND FORMATION AMONG THE INTERMEDIATE FILAMENTS DURING
CC TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.
CC -1- PTM: FLAGGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE,
CC HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM COPIES
CC OF 248 AA, WHICH ARE NOT SEPARATED BY "LARGE LINKER". THE
CC PRECURSOR IS DEPOSITED AS KERATOHYALIN GRANULES. DURING TERMINAL
CC DIFFERENTIATION IT IS DEPHOSPHORYLATED & PROTEOLYTICALLY CLEAVED.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC CC EMBL: J03458; AAT75559.1; -
CC
DR PIR: A28444; A28444.
DR MGD: MGI:95553; F19.
KW Phosphorylation; Polypeptide; Developmental protein.
FT NON_TER 1 1
SQ SEQUENCE 336 AA; 35678 MW; 259F12AD3AC0DB2D CRC64;

Query Match 25.4%; Score 64; DB 1; Length 336;
Best Local Similarity 34.5%; Pred. No. 3.5;
Matches 20; Conservative 5; Mismatches 19; Indels 14; Gaps 3;

OY 3 GHSSGSHHTTQGRSDASRGSSG-----SNSTRERTRDQEGSGDSRRHSG 48
DB 225 GHSSRHQ--ADSPRYARSGSGGGRGQSPDASGRHSNRKRRPPSPSGSDSCVHSG 280

Search completed: January 28, 2002, 07:37:50
Job time: 20 sec

```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2002, 07:37:30 : Search time 22.14 Seconds
(without alignments)
323.728 Million cell updates/sec

Title: US-09-582-711-3
Perfect score: 252
Sequence: 1 STGHSGSQSHHTTQGRSDA.....TSRETRDQEGSGDSRHS GS 49

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_rodent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	252	100.0	990	4	015206 homo sapien
2	252	100.0	1218	4	005331 homo sapien
3	218	86.5	797	4	016824 homo sapien
4	217	86.1	1084	4	001212 homo sapien
5	213	84.5	798	4	09H4U3 homo sapien
6	212	84.1	465	4	003838 homo sapien
7	207	82.1	591	4	001720 homo sapien
8	207	82.1	687	4	09H4U2 homo sapien
9	197	78.2	322	4	075370 homo sapien
10	81	32.1	727	2	053724 homo sapien
11	81	32.1	727	2	09R744 homo sapien
12	81	32.1	727	2	09R744 homo sapien
13	79	31.3	300	5	097195 leishmania
14	77.5	30.8	397	5	09V518 drosophila
15	77	30.6	343	10	081761 arabidopsis
16	76	30.2	971	5	09V086 drosophila
17	76	30.2	1913	5	09GR36 leucophaea
18	74	29.4	1151	5	09VAY4 drosophila
19	73.5	29.2	337	13	091708 xenopus lae

20	73.5	29.2	716	2	053729 staphylococ
21	73	29.0	586	4	09U089
22	73	28.0	717	4	09H307
23	73	29.0	725	11	035691 mus musculu
24	73	29.0	743	4	099738 homo sapien
25	73	29.0	1002	10	09SM75 oryza sativ
26	72.5	28.8	228	4	09NYG3 homo sapien
27	72.5	28.8	305	4	075308 homo sapien
28	72.5	28.8	305	4	015287 homo sapien
29	72.5	28.8	305	11	099M28 mus musculu
30	72	28.6	989	5	09W254 drosophila
31	71.5	28.4	448	5	09VW91 drosophila
32	71.5	28.4	828	5	09YKR7 drosophila
33	71	28.2	193	11	09D570 mus musculu
34	71	28.2	196	5	019666
35	70.5	28.0	365	4	09Y3V8 homo sapien
36	70	27.8	460	5	094504 dictyostell
37	70	27.8	811	2	09AD50 streptomyce
38	70	27.8	912	5	09V3R1 drosophila
39	69.5	27.6	200	4	003837 homo sapien
40	68.5	27.6	322	5	023796 chironomus
41	69.5	27.6	349	5	09XYH7 toxoplasma
42	69.5	27.6	471	5	09UAY0 caenorhabd
43	69.5	27.6	703	6	P79122 bos taurus
44	69.5	27.6	1217	5	017240 bombyx mori
45	69	27.4	284	6	099331 sus scrofa

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	990 AA.
015206	015206	PRELIMINARY:	PRT:	990 AA.
AC	015206			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DR	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	PROTILAGGRIN (FRAGMENT).			
GN	FLG.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=PLACENTA;			
RX	MEDLINE=91064347; PubMed=2248957;			
RA	Gan S.O., McBride O.W., Idler W.W., Markova N., Steinert P.M.;			
RT	"Organization, structure, and polymorphisms of the human profilaggrin			
RL	gene.";			
RL	Biochemistry 29:9432-9440(1990).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=PLACENTA;			
RX	MEDLINE=91255199; PubMed=2043621;			
RA	Gan S.O., McBride O.W., Idler W.W., Markova N., Steinert P.M.;			
RT	"Organization, structure, and polymorphisms of the human profilaggrin			
RL	gene.";			
RL	Biochemistry 30:5814-5814(1991).			
DR	EMBL: M60494; AAA63244.1; -;			
DR	InterPro: IPR003303; Filaaggrin.			
DR	PRINTS: PR00487; Filaaggrin.			
FT	NON_TER 990			
SO	SEQUENCE 990 AA; 106452 MW; A8396F10F6A91991 CRC64;			
Query Match				
Best Local Similarity 100.0%; Score 252; DB 4; Length 990;				
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
1 STGHSGSQSHHTTQGRSDASGSRSTRETRDQEGSGDSRHS GS 49				

Db 316 STGHSGSHSHHTTQGRSDASRGSSGSRSTRERPDQSGDGRHSGS 364

RESULT 2

005331 PRELIMINARY: PRT: 1218 AA.

AC 005331;

DT 01-NOV-1996 (TREMblrel. 01, Created)

DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)

DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)

DE FLAGGRIN (PROFLAGGRIN) (FRAGMENT).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=FORESKIN;

RA MEDLINE=93109348; PubMed=8417356;

RA Markova N.G., Marekov L.N., Chipkev C.C., Gan S.-Q., Idler W.W., Steinhert P.M.;

RT "Profilaggrin is a major epidermal calcium-binding protein.";

RL Mol. Cell. Biol. 13:613-625(1993).

CC -1- FUNCTION: AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES DISULFID-BOND FORMATION AMONGST THE INTERMEDIATE FILAMENTS DURING TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.

CC -1- PFM: FLAGGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE, HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM COPIES OF 324 AA. THE PRECURSOR IS DEPOSITED AS KERATOHALIN GRANULES DURING TERMINAL DIFFERENTIATION IT IS DEPHOSPHORYLATED & PROTEOLYTICALLY CLEAVED.

CC CC POLYMORPHISM: A NUMBER OF PROFILAGGRIN ISOFORMS HAVE BEEN FOUND WHICH DIFFER BOTH IN SEQUENCE AND IN THE NUMBER OF FLAGGRIN REPEATS.

CC EMBL: M96943; AAA36487.1; -

DR HSP; P02638; ICP.

DR InterPro: IPR002048; EF-hand.

DR InterPro: IPR003303; Flaggrin.

DR InterPro: IPR001751; S100_Cabp.

DR Pfam: PF00036; efhand; 1.

DR Pfam: PF01023; S_100; 1.

DR PRINTS: PR00487; FLAGGRIN.

DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.

DR PROSITE: PS00303; S100_CABP; 1.

KW Phosphorylation; Polyprotein; Developmental protein; Calcium-binding; Polymorphism.

FT CA_BIND 19 32 SITE 1 (BY SIMILARITY).

FT NON_TER 62 73 SITE 11 (BY SIMILARITY).

FT SEQUENCE 1218 AA; 133604 MW; EC195AD5285B19C2 CRC64;

SO

Query Match 100.0%; Score 252; DB 4; Length 1218;

Best Local Similarity 100.0%; Pred. No. 1.4e-22;

Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STGHSGSHSHHTTQGRSDASRGSSGSRSTRERPDQSGDGRHSGS 49

Db 538 STGHSGSHSHHTTQGRSDASRGSSGSRSTRERPDQSGDGRHSGS 586

RESULT 3

016824 PRELIMINARY: PRT: 797 AA.

AC 016824;

DT 01-NOV-1996 (TREMblrel. 01, Created)

DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)

DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)

DE FLAGGRIN PRECURSOR (FRAGMENT).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC MEDLINE=91064347; PubMed=2248957;

RA Gan S.-Q., McBride O.W., Idler W.W., Markova N., Steinhert P.M.;

RT "Organization, structure, and polymorphisms of the human profilaggrin gene.";

RL Biochemistry 29:9432-9440(1990).

DR EMBL: M60502; AAA63248.1; -

DR InterPro: IPR003303; Flaggrin.

DR PRINTS: PR00487; FLAGGRIN.

FT NON_TER 1 1

FT SEQUENCE 797 AA; 85176 MW; 60E6184763BDA86B CRC64;

SO

Query Match 86.5%; Score 218; DB 4; Length 797;

Best Local Similarity 85.4%; Pred. No. 1.3e-18;

Matches 41; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 2 TGHSGSHSHHTTQGRSDASRGSSGSRSTRERPDQSGDGRHSGS 49

Db 193 SGHSGSHSHHTTQGRSDASRGSSGSRSTRERPDQSGDGRHSGS 240

RESULT 4

001212 PRELIMINARY: PRT: 1084 AA.

AC 001212; 003840; -

DT 01-NOV-1996 (TREMblrel. 01, Created)

DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)

DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)

DE FLAGGRIN (FRAGMENT).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=PLACENTA;

RA MEDLINE=91064347; PubMed=2248957;

RA Gan S.-Q., McBride W.O., Idler W.W., Markova N., Steinhert P.M.;

RT "Organization, structure, and polymorphisms of the human profilaggrin gene.";

RL Biochemistry 29:9432-9440(1990).

CC -1- FUNCTION: FLAGGRIN AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES DISULFID-BOND FORMATION AMONGST THE INTERMEDIATE FILAMENTS DURING TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.

CC -1- MISCELLANEOUS: FLAGGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE, HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM COPIES OF 317 AA, WHICH ARE SEPARATED BY A SHORT LINKER SEQUENCE (PROBABLY FLYOVST). THE PRECURSOR IS DEPOSITED AS KERATOHALIN GRANULES. BY MEANS OF DEPHOSPHORYLATION AND PROTEOLYTIC CLEAVAGE FLAGGRIN IS FORMED.

CC EMBL: M60503; AAA63243.1; -

DR EMBL: M60501; AAA63243.1; JOINED.

DR InterPro: IPR003303; Flaggrin.

DR PRINTS: PR00487; FLAGGRIN.

KW Phosphorylation; Polyprotein; Developmental protein; Keratin; Intermediate filament.

FT NON_TER 1 1

FT SEQUENCE 1084 AA; 115271 MW; 80C4640B8D5A362D CRC64;

SO

Query Match 86.1%; Score 217; DB 4; Length 1084;

Best Local Similarity 83.7%; Pred. No. 2.2e-18;

Matches 41; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 STGHSGSHSHHTTQGRSDASRGSSGSRSTRERPDQSGDGRHSGS 49

Db 153 STGHSGSHSHHTTQGRSDASRGSSGSRSTRERPDQSGDGRHSGS 201

RESULT 5

```

09H4U3
ID 09H4U3 PRELIMINARY: PRT: 798 AA.
AC 09H4U3.
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Djl4N1.1.2 (PROFILAGGRIN 3' END) (FRAGMENT).
CN
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Laitid G.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL356504; CAC3171.1; -
DR InterPro; IPR003303; Filaggrin.
DR PRINTS; PR00487; FILAGGRIN.
FT NON_TER
SQ SEQUENCE 798 AA; 84773 MW; F923DDA8D1290805 CRC64;

Query Match 84.5%; Score 213; DB 4; Length 798;
Best Local Similarity 83.3%; Pred. No. 5.2e-18;
Matches 40; Conservative 4; Mismatches 4; Indels 0; Gaps 0

0Y 2 TCGSHGSHSTTTTQGRSDASRGGSGSRSTRTDQSGDGSRHSGS 49
:||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 518 SCHSSSHSHHTTSGGRSDASHGSGSRASRRTRENGSGDGSRHSGS 565

RESULT 6
003838 PRELIMINARY: PRT: 465 AA.
ID Q03838
AC Q03838;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE FILAGGRIN (PROFILAGGRIN) (FRAGMENT).
GN
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=91064347; PubMed=2248957;
RA Gan S.-O., McBride O.W., Idler W.W., Markova N., Steinhert P.M.;
RT "Organization, structure, and polymorphisms of the human profilaggrin
gene."
RL Biochemistry 29:9432-9440(1990).
RN [2]
RP REVISIONS.
RX MEDLINE=91255199; PubMed=2043621;
RA Gan S.-O., McBride O.W., Idler W.W., Markova N., Steinhert P.M.;
RT "Organization, structure, and polymorphisms of the human profilaggrin
gene."
RL Biochemistry 30:5814-5814(1991).
RN [3]
RP FUNCTION: FILAGGRIN AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND
CC PROMOTES DISULFID-BOND FORMATION AMONGST THE INTERMEDIATE
CC FILAMENTS DURING TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.
CC -1- POLYMORPHISM: A NUMBER OF PROFILAGGRIN ISOFORMS HAVE BEEN FOUND
CC WHICH DIFFER BOTH IN SEQUENCE AND IN THE NUMBER OF FILAGGRIN
CC REPEATS.
CC -1- MISCELLANEOUS: FILAGGRIN IS INITIALLY SYNTHESIZED AS A LARGE,
CC INSOLUBLE, HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM
CC COPIES OF 324 AA. THE PRECURSOR IS DEPOSITED AS KERATOHYALIN
CC GRANULES. DURING TERMINAL DIFFERENTIATION IT IS DEPHOSPHORYLATED &
CC PROTEOLYTICALLY CLEAVED.
CC EMBL: M60499; AAA63246.1; -.
DR InterPro; IPR003303; Filaggrin.

```

DR	PRINTS: PR00487; FLAGGRIN.
FT	NON_TER 1 1
FT	NON_TER 465 465
SQ	SEQUENCE 465 AA; 50280 MW; C883744C5E134097 CRC64;
Query Match	
Best Local Similarity	84.1%; Score 212; DB 4; Length 465;
Matches 39; Conservative	Pred. No. 4,1e-18; Mismatches 3; Indels 0; Gaps
Oy	2 TGHSSOHSHTTGGRPDASRGSGSRSTRETDQEGCDGSHSGS 49 : : : : : :
Db	317 TGHSSHHSHHTTGCRSDASHGTGSGRSASQTNRNEDSGDGTHSGS 364
RESULT	7
ID	001720 PRELIMINARY; PRT; 591 AA.
AC	001720.
DT	01-NOV-1996 (TREMBLrel_01, Created)
DT	01-NOV-1996 (TREMBLrel_01, Last sequence update)
DT	01-JUN-2001 (TREMBLrel_17, Last annotation update)
DE	FLAGGRIN PRECURSOR (PROFLAGGRIN) (FRAGMENT).
GN	FIG.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=PLACENTA;
RX	MEDLINE=93054736; PubMed=1429717;
RA	Presland R.B., Haydock P.V., Fleckman P., Mirunsuksiri W., Dale B.A.; "Characterization of the human epidermal profilaggrin gene. Genomic organization and identification of an S-100-like calcium binding domain at the amino terminus."
RL	J. Biol. Chem. 267:23772-23781(1992).
CC	-1 FUNCTION: AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES DISULFID-BOND FORMATION AMONGST THE INTERMEDIATE FILAMENTS DURING TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.
CC	-1 PTM: FLAGGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE, HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM COPIES OF 32A AA. THE PRECURSOR IS DEPOSITED AS KERATOHYALIN GRANULES. DURING TERMINAL DIFFERENTIATION IT IS DEPHOSPHORYLATED & PROTEOLYTICALLY CLEAVED.
CC	-1 POLYMORPHISM: A NUMBER OF PROFILAGGRIN ISOFORMS HAVE BEEN FOUND WHICH DIFFER BOTH IN SEQUENCE AND IN THE NUMBER OF FLAGGRIN REPEATS.
CC	EMBL; L01089; AAA60177.1; -
DR	EMBL; L01090; AAA60176.1; -
DR	HSSP; P02638; ICFP.
DR	MIM; 135940; -
DR	InterPro; IPRO02048; EF-hand.
DR	InterPro; IPRO03303; Flaggrin.
DR	InterPro; IPRO01751; S100_Cabp.
DR	Pfam; PF00036; efhand; 1.
DR	Pfam; PF01023; S_100; 1.
DR	PRINTS: PR00487; FLAGGRIN.
DR	PROSITE; PS00018; EF_HAND; UNKNOWN_L.
KW	PROSITE; PS00303; S100_CABP; 1.
KW	Polymorphism.
FT	PROPEP 1 293
FT	CHAIN 294 467 POTENTIAL.
FT	PROPEP 468 474 POTENTIAL.
FT	CHAIN 475 >591 FLAGGRIN.
FT	CA_BIND 19 32 SITE 1 (BY SIMILARITY).
FT	CA_BIND 62 73 SITE II (BY SIMILARITY).
FT	NON_TER 591 591
SQ	SEQUENCE 591 AA; 66366 MW; 381491625C75E369 CRC64;
Query Match	
Best Local Similarity	82.1%; Score 207; DB 4; Length 591;
Matches 79; Conservative	Pred. No. 2,1e-17;

RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhan G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*,"
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003830; AAF98818.1; -
 DR FlyBase; FBgn0033501; CG12911.
 SQ SEQUENCE 397 AA; 43204 MW; E49EA27E50CAECAC CRC64;

Query Match 30.8%; Score 77.5; DB 5; Length 397;
 Best Local Similarity 37.5%; Pred. No. 0.082;
 Matches 18; Conservative 7; Mismatches 22; Indels 1; Gaps 1;

OY 1 STGSHSGSHHTTGGRRDASRGSGSRSTRSTRETDQSGDGSRRHG 48
 DB 289 SSGSGGSGSSSSSSSSSSSSSSSGSGGSRERERE-RDRERDAQSSSYGG 335

RESULT 15

ID 081761 PRELIMINARY; PRT; 343 AA.
 AC 081761;
 DT 01-NOV-1998 (TREMBlrel. 08, Created).
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update).
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update).
 DE HYPOTHEtical 36.1 KDA PROTEIN.
 GN F1715.120 OR ATAG33930.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Vitale D., Liguori R., Argirou A., De Simone V., Hohelsel J.,
 RA Jesse T., Heljnen L., Vos P., Mewes H.W., Mayer K.F.X.,
 RA Bevan M.;
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Vitale D., Liguori R., Argirou A., De Simone V., Mewes H.W.,
 RA Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL031032; CAA19875.1; -
 DR EMBL; AL161584; CAB80110.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 343 AA; 36104 MW; 092E48BA313B242F CRC64;

Query Match 30.6%; Score 77; DB 10; Length 343;
 Best Local Similarity 42.0%; Pred. No. 0.083;
 Matches 21; Conservative 4; Mismatches 19; Indels 6; Gaps 2;

OY 5 SSGSHSH---TTTGGSDASRGSGSRSTR--ETRDQSGDGSRRHG 48
 DB 89 SGNHSSGTGCTHNGHSSGNHSSATGCTHNGHTRSTGSHSSGNNGSRHNG 138

Search completed: January 28, 2002, 07:38:39
 Job time: 69 sec